DOCUMENT RESUME

ED 057 748 HE 002 756

AUTHOR	Hildebrand, Milton; And Others
TITLE	Evaluating University Teaching.
INSTITUTION	California Univ., Berkeley. Center for Research and Development in Higher Education.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Cooperative Research Program.
BUREAU NO	BR-5-0248
PUB DATE	71
GRANT	OEG-6-10-106
NOTE	57p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Educational Development; *Educational Improvement; Faculty Evaluation; *Higher Education; *Instructional Improvement; *Student Opinion; *Teaching Methods

ABSTRACT

This study defines and describes effective teaching so that instructors can be helped to improve and so that graduate students can be better prepared for the teaching function of academic life. The means devised to accomplish this is a series of evaluations in the form of questionnaires that involved more than 1,600 students and faculty over the 3-year study period. The report contains chapters concerning the development of the teacher description scales, the ratings of teachers related to characteristics of courses and students, and the results of the evaluations. A final chapter deals with finding a more valid, reliable, and effective means of incorporating the evaluation of teaching into advancement procedures. (HS)

ED057748

HE
007

EVALUATING UNIVERSITY TEACHING

MILTON HILDEBRAND
ROBERT C. WILSON
EVELYN R. DIENST

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

HEW 2 T56

CENTER FOR RESEARCH AND DEVELOPMENT IN HIGHER EDUCATION
UNIVERSITY OF CALIFORNIA, BERKELEY

1

CENTER FOR RESEARCH AND DEVELOPMENT IN HIGHER EDUCATION
UNIVERSITY OF CALIFORNIA, BERKELEY

The Center for Research and Development in Higher Education is engaged in research designed to assist individuals and organizations responsible for American higher education to improve the quality, efficiency, and availability of education beyond the high school. In the pursuit of these objectives, the Center conducts studies which: 1) use the theories and methodologies of the behavioral sciences; 2) seek to discover and to disseminate new perspectives on educational issues and new solutions to educational problems; 3) seek to add substantially to the descriptive and analytical literature on colleges and universities; 4) contribute to the systematic knowledge of several of the behavioral sciences, notably psychology, sociology, economics, and political science; and 5) provide models of research and development activities for colleges and universities planning and pursuing their own programs in institutional research.

The research reported herein was supported by Grant No. OE-6-10-106, Project No. 5-0248-2-8A, with the Office of Education, U.S. Department of Health, Education, and Welfare, under the provision of the Cooperative Research Program. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

ED057748

EVALUATING UNIVERSITY TEACHING

MILTON HILDEBRAND
ROBERT C. WILSON
EVELYN R. DIENST

A Handbook published by

CENTER FOR RESEARCH AND DEVELOPMENT IN HIGHER EDUCATION
UNIVERSITY OF CALIFORNA, BERKELEY 1971

S3

Contents

FOREWORD 1

DEVELOPMENT OF TEACHER DESCRIPTION SCALES 4

- Collection of data 4
- Identification of effective teachers 5
- Teaching characterized–by students 7
- Teaching characterized–by colleagues 13
- Components of effective teaching 16
- Usefulness of the scales 22

RATINGS OF TEACHERS RELATED TO CHARACTERISTICS OF COURSES AND STUDENTS 27

- Courses and students 27
- Goals of students 29
- Matching students with teachers 33

EVALUATIONS DISCUSSED 34

- What is effective teaching? 34
- Comparison of evaluations by students and colleagues 36
- Sample size and norms 38
- A potential weakness in the use of student evaluations 39
- Alternative student evaluation forms 39

THE PRINCIPAL RESULTS 41

SOME IMPLICATIONS 43

PLANNING FOR PROGRAMS OF TEACHER EVALUATION 45

REFERENCES 49

Evaluating University Teaching

FOREWORD

One major aim of this study is to define and describe effective teaching so that instructors can be helped to improve, and graduate students can be better prepared for the teaching function of academic life. Articles allegedly describing good teaching are numerous, and many are sound, but most either largely represent the subjective judgment of individuals and committees, or are based on studies using small samples in restricted circumstances. Reliable characterization of effective teaching is needed.

The other major aim is to find more valid, reliable, and effective means of incorporating the evaluation of teaching into advancement procedures. We believe this to be the most important single requirement for the improvement of university teaching; the incentive thereby provided will encourage instructors to devote the study, time, and effort necessary to do their best, and the status of teaching will increase.

Because procedures for evaluating teaching have been largely unstandardized and untested, research productivity usually

has outweighed quality of teaching as a criterion for advancement. Yet, in a recent survey (Wilson, Gaff, & Bavry, 1970) of 1000 faculty members at six diverse colleges and universities, 92 percent stated that teaching effectiveness should be *quite important* or *very important* as a criterion for advancement, whereas only 38 percent of the sample stated that effectiveness as a teacher actually is either *quite important* or *very important*. No less than 72 percent of the respondents felt that their campuses should have a formal procedure for evaluating teaching.

At most colleges and universities, the dossier furnished by the department chairman to support promotion has been of the utmost importance (Gustad, 1961), yet there are inherent weaknesses in a system that places great weight on evaluations of teaching as traditionally prepared by chairmen (or deans): A chairman may himself be doubtfully qualified as a judge of teaching, and opinions solicited from his staff may be biased or not constitute an adequate sample, and often are in part second hand. Most available measures of involvement in teaching (such as number of courses taught, enrollments, number of advisees) do not necessarily correlate with quality of instruction. Classroom visitations are resisted or resented by most teachers, and hence are seldom made, although they are considered by many administrators to be the most important element in evaluation. In any event, if a department is large, the chairman cannot visit any class more than once or twice, which is enough to judge certain elements of effective teaching, but insufficient to make a comprehensive judgment. Classroom instruction, after all, is only part of the teaching function.

We believe that promotion letters cannot be improved sufficiently to achieve our objective unless new procedures assure that they include more thorough, more objective, and more comparable evaluations of teaching than have been usual in the past.

This three-year study, which involved more than 1600 students and faculty was recommended in 1966 by an ad hoc Committee on Teaching of the University of California, Davis. Funds provided by the president of the university and the chancellor of the Davis campus were supplemented by the Center for Research and Development in Higher Education, University of California, Berkeley. Milton Hildebrand, Professor of Zoology at Davis, and Robert C. Wilson, Research Psychologist at the Center, were co-directors of the study. Hildebrand posed the problems and actively participated in the interpretation of the results and writing of this report; Wilson designed the study, supervised its conduct, and edited the final report. Evelyn R. Dienst assisted in all phases of the study. Nancy Watson made many valuable contributions, particularly in data analysis. We thank the Faculty Advisory Committee, which reviewed the research plan, questionnaires, and drafts of the report; members were H. L. Alder, R. W. Hoermann, R. M. Johnson, M. P. Oettinger, M. Regan, G. D. Yonge, and P. E. Zinner. We also thank Wilbert J. McKeachie, Kenneth E. Eble, and the many other persons who reviewed a draft of the report. Harriet Renaud provided invaluable aid and professional wisdom in editing the report and supervising its production. Thanks are also due Patsy Babbitt, of the Center's Development and Dissemination section, for her conscientious typing of the manuscript, and her creative attention to the details of its final production.

Copies of both a short-form and medium-length form for obtaining student and colleague descriptions of teachers are available from the Center for Research and Development in Higher Education, University of California, Berkeley. Implementation of the suggested teacher evaluation procedures is considered one of the Center's development obligations. Permission for use of the forms, and assistance in initiating and carrying out programs for the improvement of teaching are available upon request.

8

7

DEVELOPMENT OF TEACHER-DESCRIPTION SCALES

Collection of Data

Three questionnaires were distributed in May 1967, and one in May 1968. Of the random sample of all students asked to complete the first questionnaire, 278 undergraduate and 60 graduate students responded (4 percent of the student body and 38 percent of those approached). The respondents were evenly divided between the sexes, did not differ significantly from the population in distribution by class level or major area, and had a mean overall grade point average identical with the grade point average of the population for that quarter. Respondents supplied biographical information and their academic backgrounds, answered questions about their college goals and the objectives they valued in teaching, and described the teaching of those identified by them as the *best* instructors and *worst* instructors they had had in the previous year. Assurance was given that the identity of teachers would be kept in strict confidence.

The second questionnaire was returned by 119 of the faculty (54 percent of the random sample approached and 21 percent of the resident teaching faculty). Respondents were asked to identify a *best* and a *worst* teacher among their colleagues and to answer, for each, questions about teaching activities observed outside the classroom, about in-class behavior, and about the presentation of talks and seminars.

The third questionnaire, dealing with the distribution of time among various academic pursuits, was returned by 162 members of the faculty who had not been asked to complete the previous questionnaire (80 percent of the random sample approached and 29 percent of the resident teaching faculty).

Lastly, as a follow-up and validation study, a fourth questionnaire was distributed in 1968 to all students in 51 classes. The classes selected included, in about equal numbers, those of instructors identified in 1967 as *best* teachers by three or more students or colleagues, those of instructors identified as *worst* teachers, and those of instructors not previously identified as either *best* or *worst,* and presumed to be teachers of intermediate effectiveness. The 1015 respondents provided biographical data and answered questions about their college goals, various objectives of teaching, and the teaching of the given instructor. Ratings of the overall effectiveness of the teachers were also secured.

Identification of Effective Teachers

One of the questions most frequently raised about teaching effectiveness is whether the various segments of the academic community agree in their identifications of effective and ineffective teachers. To answer this question, instructors were identified who received either three or more nominations as *best* teachers or three or more nominations as *worst* teachers from the respondents to the 1967 survey. In earlier, unpublished study done at the same campus by Regan and Yonge, 57 of the same teachers were named by students as being particularly excellent or poor. Table 1 shows the very high degree of agreement between the two surveys: The chi square value indicates a level of significance of $p < .0005$ (that is, fewer than 5 chances in 10,000 that the observed result is fortuitous).

This result indicates that the two groups of students probably used closely similar criteria. Since the Regan and Yonge study had a 90 percent return, this is considered indirect evidence that self-selection did not introduce significant bias into the present respondents' designations of *best* and *worst* teachers.

TABLE 1

AGREEMENT BETWEEN NOMINATIONS FOR *BEST* AND *WORST* TEACHERS BY TWO STUDENT SAMPLES

1967 Student nominations	1963-1966 Student nominations: Best	1963-1966 Student nominations: Worst	
Best	26	3	N = 57
Worst	4	24	chi square = 29.1 $p < .0005$

Further, in the 1968 survey, ratings were given instructors by all students of 15 instructors named in 1967 by three or more students as *best* teachers (or by a margin of three *best* over *worst* nominations if the teacher was given both kinds of ratings), all students of 18 instructors named previously as *worst* teachers, and all students of 18 instructors not previously nominated as either *best* or *worst*. Ratings were along a seven-point continuum from *Among the very worst* to *Among the very best*. Differences between the mean scores for *best, not nominated*, and *worst* teachers of the previous year were all significant well below the .01 level. Mean scores for *best, not nominated,* and *worst* teachers were respectively 6.16 ($s = 1.02$, $N = 573$), 5.28 ($s = 1.39$, $N = 297$), and 4.58 ($s = 1.59$, $N = 283$). For the difference between *best* and *worst*, $p < .0005$, for the difference between *best* and *not nominated*, $p < .005$, and for the difference between *not nominated* and *worst*, $p < .01$. ($N > 1015$ because responses are included that were eliminated from subsequent analysis.)

Finally, each of 119 faculty respondents identified colleagues they considered outstanding teachers and those they

considered poor teachers. Of those named, 66 were common to the choices of the 1967 student sample. Table 2 shows the very high agreement between the two groups; again, $p < .0005$.

TABLE 2

AGREEMENT BETWEEN NOMINATIONS FOR *BEST* AND *WORST* TEACHERS BY 1967 STUDENT SAMPLE AND A SAMPLE OF THE FACULTY

Faculty nominations	Best	Worst	
Best	37	8	N = 66
Worst	2	19	chi square = 31.3
			$p < .0005$

Having learned that there is excellent agreement among students, and between faculty and students, about the effectiveness of given teachers, the next step was to characterize effective teaching.

Teaching Characterized–by Students

The student respondents to the 1967 survey indicated whether each of 158 descriptions of aspects of teaching (shortened from 236 items after analysis of a pretest taken by 44 students) was characteristic for the instructors they named as their *best* and *worst* teachers of the year. Possible answers were *Yes, No,* and *Does not apply or don't know.* The respondents to the 1968 survey stated whether most of the same items (and some new ones) were descriptive of their teachers, this time using a four-point scale ranging from *Not at all descriptive* to *Very descriptive.* Items were drawn from the experience of the research staff and the faculty advisory committee, and from studies by other investigators

(Cosgrove, 1959; Crannel, 1953; Gibb, 1955; Guthrie, 1954; Hayes, 1963; Hodgson, 1958; Isaacson, 1964; Lacognata, 1964; Rezler, 1965; Ryans, 1960; Solomon, 1966; Solomon et al., 1964).

Table 3 lists 85 of the 158 items to which at least 75 percent of respondents could answer *Yes* or *No*, and which discriminate between *best* and *worst* teachers with the very high significance level of $p < .001$. For easier tabulation in the text, many of the items have been somewhat condensed.

TABLE 3

CHARACTERIZATION OF EFFECTIVE TEACHERS–BY STUDENTS

Characteristics of a Majority of *Best* Teachers and a Minority of *Worst*

Course Content and Presentation

†	* 1.	Contrasts implications of various theories
	2.	Presents origins of ideas and concepts
	* 3.	Presents facts and concepts from related fields
	4.	Talks about research he has done himself
	5.	Emphasizes ways of solving problems rather than solutions
	6.	Discusses practical applications
	7.	Explains his actions, decisions, and selection of topics
†	8.	Seems well read beyond the subject he teaches
	* 9.	Is an excellent public speaker
†	10.	Speaks clearly
	*11.	Explains clearly
	12.	Gives lectures that are easy to outline
	13.	Reads lectures or stays close to notes (Negative)
	14.	Assigns text, but lectures include other topics
	*15.	Makes difficult topics easy to understand
	16.	Summarizes major points
	17.	States objectives for each class session
	18.	Identifies what he considers important
	*19.	Shows interest and concern in quality of his teaching
	20.	Gives examinations requiring creative, original thinking

21. Gives examinations having instructional value
22. Gives examinations requiring chiefly recall of facts (Negative)
23. Gives interesting and stimulating assignments
24. Stresses the aesthetic and emotional value of the subject
* 25. Is a dynamic and energetic person
†* 26. Seems to enjoy teaching
† 27. Is enthusiastic about his subject
† 28. Seems to have self-confidence
29. Varies the speed and tone of his voice
30. Has a sense of humor

Relations with Students

31. Is careful and precise in answering questions
† 32. Explains his own criticisms
33. Encourages class discussion
* 34. Invites students to share their knowledge and experiences
* 35. Clarifies thinking by identifying reasons for questions
* 36. Invites criticism of his own ideas
†* 37. Knows if the class is understanding him or not
38. Knows when students are bored or confused
39. Has students apply concepts to demonstrate understanding
†* 40. Keeps well informed about progress of class
41. Anticipates difficulties and prepares students beforehand
42. Has definite plan, yet uses material introduced by students
43. Provides time for discussion and questions
* 44. Is sensitive to student's desire to ask a question
45. Encourages students to speak out in lecture or discussion
† 46. Quickly grasps what a student is asking or telling him
47. Restates questions or comments to clarify for entire class
48. Asks others to comment on one student's contribution
49. Compliments students for raising good points
50. Doesn't fully answer questions (Negative)
51. Determines if one student's problem is common to others
52. Reminds students to see him if having difficulty
53. Informs students of coming campus events related to course
54. Encourages students to express feelings and opinions

SD

55. Relates class topics to students' lives and experiences
† 56. Has a genuine interest in students
57. Relates to students as individuals
58. Recognizes and greets students out of class
*59. Is valued for advice not directly related to the course
60. Treats students as equals

Characteristics of a Majority of *Best* and *Worst* Teachers, But More Typical of *Best*

61. Discusses points of view other than his own
62. Discusses recent developments in the field
63. Gives references for the more interesting and involved points
64. Emphasizes conceptual understanding
65. Disagrees with some ideas in textbook and other readings
66. Stresses rational and intellectual aspects of the subject
67. Stresses general concepts and ideas
68. Seems to have a serious commitment to his field
69. Is well prepared
70. Gives examinations stressing conceptual understanding
71. Gives examinations requiring synthesis of various parts of course
72. Gives examinations permitting students to show understanding
73. Is friendly toward students
74. Is accessible to students out of class
75. Respects students as persons
76. Is always courteous to students
77. Gives personal help to students having difficulty with course
78. Has an interesting style of presentation

Results Typical of Taking a Course from a *Best* Teacher and not from a *Worst*

†*79. Have developed increased appreciation for the subject
†*80. Have learned new ways to evaluate problems
81. Have worked harder than in most other courses

82. Know how to find more information on the subject
83. Have studied a topic from the course on own initiative
84. Plan to take more courses on the subject
85. Have gained self-knowledge

*Descriptive of 75% or more of *best* teachers and 25% or less of *worst* teachers
†Descriptive of 95% or more of *best* teachers and 45% or less of *worst* teachers
Items not listed in rank order

While this table goes far toward providing a description of fine teaching, the included items are not equally useful for making comparative evaluations of teaching. Because students and colleagues both tend to rate instructors generously (Gowan & Payne, 1962; Kent, 1967; Weaver, 1960) items that discriminate at the top are particularly useful. When teachers in general are rated on selected items, it is desirable that the distributions of scores not be skewed so that there are many more high than low scores. Items 1 through 60 meet this requirement better than the remaining items. Asterisks and daggers mark the most discriminating items, with those marked by asterisks also providing the least skewed distributions of scores.

Some items (numbers 61 through 78 of Table 3 are characteristic of a majority of both *best* and *worst* teachers, although sufficiently more typical of *best* teachers to discriminate at below the .001 level of significance. If teachers in general were rated on such items, one would expect the distributions to be markedly skewed: If an item were not descriptive of a given teacher, his teaching would probably not be effective in that regard, but if the item were descriptive, his teaching might still be relatively ineffective. (Examination of the items suggests that even our *worst* teachers are competent in many respects.) To use such items for evaluation is equivalent to giving an easy quiz to a class of variable but generally high excellence: All students earn 100 percent scores except the few already known to be at the foot of the class. A

department chairman who wished to write nice things in a promotion letter about a relatively mediocre teacher could probably select several such items.

A smaller category comprises items (not included in Table 3) that are characteristic of a minority of *best* and *worst* teachers, but less so of *best* teachers to the extent that $p < .001$. Examples are: *Has distracting mannerisms. Emphasizes grades. Gives ambiguous examinations.*

Nondiscriminating items should be excluded from evaluation forms (although they may be useful for other purposes, such as the selection of teachers by students). Noteworthy among items found not to distinguish *best* from *worst* teachers, even at the comparatively low .05 level of significance, were: *Gives difficult examinations. Gives difficult assignments. Spends much of his time on research or projects other than teaching. Grades leniently. Grades subjectively.* These responses, and those to numbers 5, 20, 39, 64, 66, 67, 71, 80, 83, and particularly 81, strongly indicate that students do not equate *best* teachers with *easy* teachers.

Questions to which many students are unable to reply are of limited value for evaluating teachers, particularly when classes are small. The following are representative of items that discriminate *best* from *worst* teachers, but to which at least 25 percent of the respondents could not reply: *Is always in his office during scheduled office hours. Puts me at ease when I visit him. Is involved in campus activities that affect students. Learns students' names promptly. Is well known in his field. Spends extra time with students having difficulty.*

Some items (4, 13, 14, 24, 30, 43, 48, 55, 63, 65, 76, and 85) discriminate *best* from *worst* teachers if ratings are by undergraduate students, but not if ratings are by graduate students.

The difference probably results from the nature of graduate instruction and the greater professional orientation and self-motivation on the part of graduate students.

Teaching Characterized–by Colleagues

For colleagues named as the most and least effective teachers known to them, 119 of the faculty respondents indicated whether each of 103 descriptions of aspects of teaching and other academic activities was characteristic. Answers were *Yes, No,* and *Does not apply or don't know.* Table 4, which supplements Table 3 in characterizing good teachers, lists 54 items to which at least 66 percent of respondents answered *Yes* or *No*, and which discriminated between *best* and *worst* teachers with a significance level of $p < .001$.

TABLE 4

CHARACTERIZATION OF EFFECTIVE TEACHERS–BY COLLEAGUES

Characteristics of a Majority of *Best* Teachers and a Minority of *Worst*

1. Does original and creative work
2. Expresses interest in the research of his colleagues
3. Gives many papers at conferences
4. Has done work to which I refer in teaching
5. Has been consulted by me about my research
6. Has been consulted by me about problems in his field
7. Discusses students' work with colleagues
† 8. Spends much time planning and preparing for his teaching
9. Seems well read beyond the subject he teaches
10. Is sought by others for advice on research
† 11. Can suggest reading in any area of his general field
12. Is sought by colleagues for advice on academic matters
13. Encourages students to talk with him on matters of concern
14. Is involved in campus activities that affect students
15. Attends many lectures and other events on campus
16. Enjoys controversy in discussion and may provoke opposing views

† 17. Comes to departmental or committee meetings well prepared
18. Meets with students informally out of class
19. Meets with students out of regular office hours
20. Encourages students to talk with him on matters of concern
† 21. Seems to have a congenial relationship with students
† 22. Seems to have a genuine interest in his students
*23. Seeks advice from others about the courses he teaches
† 24. Discusses teaching in general with colleagues
25. Does not seek close friendships with colleagues (Negative)
26. Is someone with whom I have discussed my teaching
27. Is interested in, and informed about, the work of colleagues
28. Expresses interest and concern about the quality of his teaching
† 29. Seems to enjoy teaching

Further Characterization if Speech or Seminar was Attended

† 30. Gives a well organized presentation
*31. Is an excellent public speaker
32. Summarizes major points at the end of the presentation
*33. Uses wit and humor effectively
† 34. Uses well chosen examples to clarify points
† 35. Communicates self-confidence

Further Characterization if Classroom Teaching was Attended

36. Encourages students to express feelings and opinions
*37. Clarifies thinking by identifying reasons for questions
38. Presents facts and concepts from related fields
*39. Anticipates difficulties and prepares students beforehand
† 40. Quickly grasps what a student is asking or telling him
† 41. Is careful and precise in answering questions
42. Presents origins of ideas and concepts
† 43. Emphasizes ways of solving problems rather than solutions

Characteristics of a Majority of *Best* and *Worst* Teachers, But More Typical of *Best*

44. Invites discussion of points he raises
45. Is careful and precise in answering questions

46. Keeps current with developments in his field
47. Has talked with me about his research
48. Knows about developments in fields other than his own
49. Has a congenial relationship with colleagues
50. Is conscientious about keeping appointments with students
51. Recognizes and greets students out of class
52. Is enthusiastic about his subject
53. Does work that receives serious attention from others
54. Corresponds with others about his research

*Descriptive of 25% or more of *best* teachers and 25% or less of *worst* teachers

†Descriptive of 95% or more of *best* teachers and 45% or less of *worst* teachers

Items not listed in rank order

The item, *Publishes frequently,* is discriminating for *best* teachers at the .05 significance level. Noteworthy among items found not to be discriminating were: *Spends much of his time on research or projects other than teaching. Attends faculty social functions. Expresses concern about pressures to publish.*

Of the numerous items to which more than one-third of our colleague respondents replied *Does not apply or don't know,* most related to instructor-student interaction.

As another part of the study, a random sample of 162 of the faculty was asked to state how often various functions of teaching, research, university and community service, consultation, and related academic pursuits had been performed in stated time periods. Of all respondents, 38 had been named as *best* teachers and 32 as *worst* teachers by students or colleagues on the independent surveys already described. When the self-descriptions of the *best* and *worst* teachers were compared, remarkably little difference was found. Only two of the 143 items, *Met informally*

with students outside of class or office, and *Talked with a colleague about my research,* discriminated between effective and ineffective teachers below the .05 level of significance. None of the other comparisons was found to be statistically significant. Examples of nondiscriminating items are: *Reviewed lecture notes. Revised a lecture. Prepared demonstration material for a class. Did background reading for a course. Graded examination papers. Helped students with individual projects.* At least within the limits of discrimination established here, the more and less effective teachers at the campus studied do the same general things with their time. Involvement with teaching on the part of candidates for promotion is a proper consideration in a recommendation report, but the mere performance of activities associated with teaching evidently does not of itself assure that the instruction is effective.

Together, the items in Tables 3 and 4 give a picture of good teaching as defined by students and colleagues. But since the list of items is long and miscellaneous in character, and does not fully characterize effective teaching in a conceptual manner, further analysis is necessary.

Components of Effective Teaching

Many researchers (among them Bendig, 1953; Coffman, 1954; Cosgrove, 1959; Crannel, 1953; Estrin, 1965; French, 1957; Garverick & Carter, 1962; Gibb, 1955; Isaacson et al., 1964; Remmers & Baker, 1952; Solomon, 1966; Solomon et al., 1964; and Wherry, 1950) have identified basic components, dimensions, or scales of effective teaching by sorting individual items describing aspects of effective teaching into related groups. Teacher-rating forms developed by students commonly do the same. Scales have been determined by subjective examination of a list of items, or by factor analysis, (which mathematically establishes the tendency of responses to the various items to associate in clusters). The

number of scales developed in these studies ranges from four to 13, with four or five particular scales *(knowledge, presentation, relation with students, enthusiasm)* appearing rather consistently, even though the terminology differs. The scales developed in this study are generally consistent with those of previous studies.

Scales characterizing effective teaching as perceived by students were established by factor analysis of 91 items describing the teaching of 338 teachers identified as *best* by respondents to the 1967 survey. (Items were eliminated from the original list of 158 if: they did not discriminate between *best* and *worst* teachers at the .001 level; 25 percent or more of respondents could not reply *Yes* or *No* to them; they were descriptive of virtually all *best* teachers, of few *best* or *worst* teachers, or of most *best* and *worst* teachers; or if they were applicable only to small classes, or related to examinations and assignments.) The method used was a principal-components analysis with a varimax rotation (Kaiser, 1958).

After several analyses, a five-factor solution was selected as giving the maximum number of distinct and interpretable components of effective teaching. Items having factor coefficients (which show the tendency of an item to be associated with a particular scale) greater than .40 were retained and analyzed further by pre-set cluster analysis (Tryon & Bailey, 1966) to determine the consistency and reliability of the scales and their intercorrelations, the highest being 3 with 4, .38; and 1 with 3, .32. The items were then re-analyzed with data from our 1968 validation survey. The five scales held together very well; the alpha reliabilities (showing internal consistency) ranged from .80 to .89. (Alpha reliabilities for the data from the 1967 survey ranged from .58 to .76, these values being lower because only *best* teachers were included in that analysis.)

Table 5 presents the five scales and the included items, none of which appears in more than one scale. The factor coefficients from the 1968 survey are listed. The 1967 values are similar; the 1968 values are shown because several new items had been added. Conceptual interpretations of the scales are:

Scale 1, *Analytic/Synthetic Approach*, relates to scholarship, with emphasis on breadth, analytic ability, and conceptual understanding.

Scale 2, *Organization/Clarity*, relates to skill at presentation, but is subject-related, not student-related, and not concerned merely with rhetorical skill.

Scale 3, *Instructor-Group Interaction*, relates to rapport with the class as a whole, sensitivity to class response, and skill at securing active class participation.

Scale 4, *Instructor-Individual Student Interaction*, relates to mutual respect and rapport between the instructor and the individual student.

Scale 5, *Dynamism/Enthusiasm* relates to the flair and infectious enthusiasm that comes with confidence, excitement for the subject, and pleasure in teaching.

TABLE 5

COMPONENTS OF EFFECTIVE TEACHING AS PERCEIVED BY STUDENTS*

Scale 1. Analytic/Synthetic Approach		Factor coefficient
1.	Discusses points of view other than his own	.70
2.	Contrasts implications of various theories	.66
3.	Discusses recent developments in the field	.64
4.	Presents origins of ideas and concepts	.60
5.	Gives references for more interesting and involved points	.53
6.	Presents facts and concepts from related fields	.53
7.	Emphasizes conceptual understanding	.46

Scale 2. Organization/Clarity		Factor coefficient
8.	Explains clearly	.78
9.	Is well prepared	.63
10.	Gives lectures that are easy to outline	.62
11.	Is careful and precise in answering questions	.61
12.	Summarizes major points	.51
13.	States objectives for each class session	.50
14.	Identifies what he considers important	.47
Scale 3. Instructor-Group Interaction		
15.	Encourages class discussion	.70
16.	Invites students to share their knowledge and experiences	.65
17.	Clarifies thinking by identifying reasons for questions	.64
18.	Invites criticism of his own ideas	.62
19.	Knows if the class is understanding him or not	.58
20.	Knows when students are bored or confused	.57
21.	Has interest and concern in the quality of his teaching	.48
22.	Has students apply concepts to demonstrate understanding	.43
Scale 4. Instructor-Individual Student Interaction		
23.	Has a genuine interest in students	.74
24.	Is friendly toward students	.71
25.	Relates to students as individuals	.69
26.	Recognizes and greets students out of class	.68
27.	Is accessible to students out of class	.65
28.	Is valued for advice not directly related to the course	.64
29.	Respects students as persons	.60
Scale 5. Dynamism/Enthusiasm		
30.	Is a dynamic and energetic person	.80
31.	Has an interesting style of presentation	.76
32.	Seems to enjoy teaching	.74
33.	Is enthusiastic about his subject	.65
34.	Seems to have self-confidence	.64
35.	Varies the speed and tone of his voice	.63
36.	Has a sense of humor	.53

*Based on 1968 survey. N = 1015

Responses describing the performance of *worst* teachers were also subjected to factor analysis, but the results showed less consistent relationships than they did for *best* teachers. Ineffective teachers thus were described by a lack of attributes associated with effective teaching, rather than by characteristics associated with poor teaching.

Scales for characterizing effective teachers by colleagues were developed by factor analysis of 67 items which described the behavior of 84 *best* teachers identified by 119 members of the faculty. Items requiring attendance of the respondent at classroom instruction and at lectures or seminars for colleagues of the identified teacher (numbers 30 through 45 of Table 4) were not factored because many colleagues (51 percent and 17 percent, respectively) had not observed those activities. Items also were excluded if not discriminating at the $p < .001$ level, and if more than 33 percent of respondents checked *Does not apply or don't know.*

Five scales were established by the same method of factor analysis as for the student data. The factor coefficients of the included items are listed in Table 6. The alpha reliabilities ranged from .65 to .86. Intercorrelations between the scales are generally low or negligible, the highest intercorrelations being 1 with 2, .41; and 3 with 4, .39. Conceptual interpretations of the scales are indicated by the headings assigned to them:

Scale 1. Research Activity and Recognition

Scale 2. Intellectual Breadth

Scale 3. Participation in the Academic Community

Scale 4. Relations with Students

Scale 5. Concern for Teaching

TABLE 6

COMPONENTS OF THE ACTIVITIES OF EFFECTIVE TEACHERS*
AS PERCEIVED BY COLLEAGUES

Scale 1. Research Activity and Recognition	Factor coefficient
1. Does work that receives serious attention from others	.69
2. Corresponds with others about his research	.69
3. Does original and creative work	.64
4. Expresses interest in the research of his colleagues	.55
5. Gives many papers at conferences	.55
6. Keeps current with developments in his field	.49
7. Has done work to which I refer in teaching	.48
8. Has talked with me about his research	.38
Scale 2. Intellectual Breadth	
9. Seems well read beyond the subject he teaches	.66
10. Is sought by others for advice on research	.60
11. Can suggest reading in any area of his general field	.59
12. Knows about developments in fields other than his own	.51
13. Is sought by colleagues for advice on academic matters	.43
Scale 3. Participation in the Academic Community	
14. Encourages students to talk with him on matters of concern	.60
15. Is involved in campus activities that affect students	.58
16. Attends many lectures and other events on campus	.47
17. Has a congenial relationship with colleagues	.39
Scale 4. Relations with Students	
18. Meets with students informally out of class	.58
19. Is conscientious about keeping appointments with students	.57
20. Meets with students out of regular office hours	.57
21. Encourages students to talk with him on matters of concern	.55
22. Recognizes and greets students out of class	.37
Scale 5. Concern for Teaching	
23. Seeks advice from others about the courses he teaches	.70
24. Discusses teaching in general with colleagues	.60

25

		Factor coefficient
25.	Does not seek close friendships with colleagues (Negative)	-.47
26.	Is someone with whom I have discussed my teaching	.45
27.	Is interested in and informed about the work of colleagues	.44
28.	Expresses interest and concern about the quality of his teaching	.40

*Based on 1967 survey. N = 119

Usefulness of the Scales

The scales derived from the characterization of effective teaching by students provide a means for conceptualizing the components of such teaching. Having been developed from items to which most students of a large random sample could respond, the student scales are applicable to most kinds of university-level teaching. The scales focus attention on the major factors to consider either in teaching or in the evaluation of teaching. Many of the rating forms used on various campuses omit items relating to one or more of the important components of teaching and thus fail in this respect.

To learn if an effective short evaluation form could be developed, a summary description of each of the student scales derived from the 1967 survey was written, to express the component of effective teaching defined by the items in each scale. The 1968 survey then asked respondents to rate their teachers on each of these five descriptions, and also repeated the full set of original items from which the scales had been established. Correlations of mean scores on the summary descriptions with mean scores on the full list of respective items (N = 51) were very high (coefficients ranging from .88 to .96). Thus, a short-form rating instrument was established that is quickly answered, yet is objectively known to be broad, balanced, and highly discriminating between effective and ineffective teachers.

The five recommended summary descriptions listed below have been modified somewhat from those used in the 1968 survey

to emphasize the items found most discriminating and to give less emphasis to items which, even though discriminating, are characteristic of both *best* and *worst* teachers. Since respondents tend to use the upper part of a rating scale, a seven-point continuum is suggested, ranging from *Not at all descriptive* to *Very descriptive* because such a continuum provides more discrimination than a five-point one at the high end of the scale.

1. *Has command of the subject, presents material in an analytic way, contrasts various points of view, discusses current developments, and relates topics to other areas of knowledge.*

2. *Makes himself clear, states objectives, summarizes major points, presents material in an organized manner, and provides emphasis.*

3. *Is sensitive to the response of the class, encourages student participation, and welcomes questions and discussion.*

4. *Is available to and friendly towards students, is interested in students as individuals, is himself respected as a person, and is valued for advice not directly related to the course.*

5. *Enjoys teaching, is enthusiastic about his subject, makes the course exciting, and has self-confidence.*

Respondents to the 1968 student survey made a single overall rating of the effectiveness of their teachers on a continuum of 1 to 7. Table 7 shows the correlations between the overall rating of effectiveness and the five separate summary descriptions. Scale 5, *Dynamism/Enthusiasm*, is the most highly related to ratings of overall effectiveness, and Scale 2, *Organization/Clarity*, is second highest. For all the correlations, $p < .001$.

TABLE 7

CORRELATIONS BETWEEN STUDENT RATINGS OF OVERALL EFFECTIVENESS

	Component	Correlation with overall rating
1.	Analytic/Synthetic Approach	.60
2.	Organization/Clarity	*.74*
3.	Instructor-Group Interaction	.59
4.	Instructor-Individual Student Interaction	.63
5.	Dynamism/Enthusiasm	*.83*

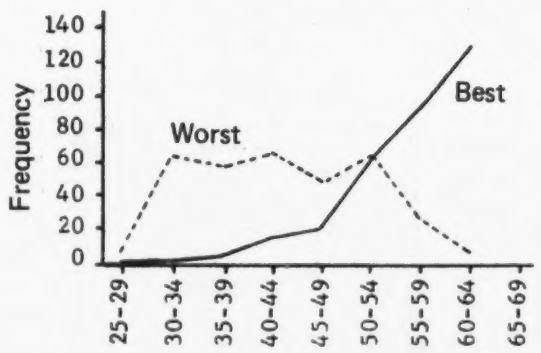
Correlations > .70 = high (italicized); .70 to .40 = moderate. N = 51

The usefulness of the five scales for discriminating *best* from *worst* teachers is shown in another way. Each teacher named in the 1967 student survey was given a score for each scale based on the total number of items students listed as descriptive of his performance. The scores for each scale were then converted so that the mean score for all teachers is 50 and the standard deviation is 10. Table 8 shows frequency distributions for the converted scores of *best* and *worst* teachers.

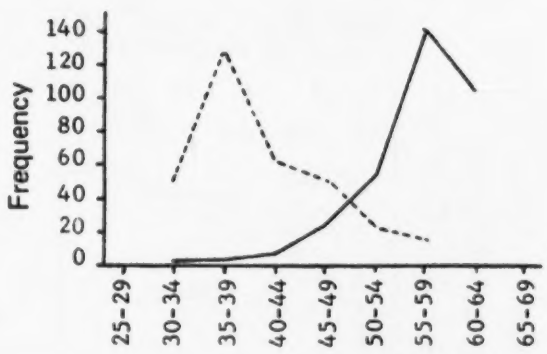
Similarly, Table 9 presents the percentages of *best* and *worst* teachers that fall within each range of the converted scores. These percentages can be interpreted as the probabilities that any teacher with a given score would be nominated by students as a *best* or a *worst* teacher.

TABLE 8

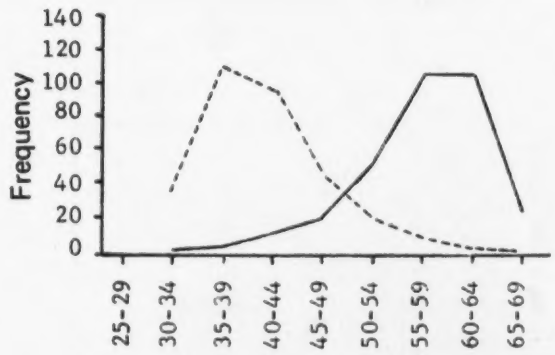
FREQUENCY DISTRIBUTIONS OF CONVERTED SCORES ($\bar{x}$ = 50, s = 10) OF 338 *BEST* AND 338 *WORST* TEACHERS ON FIVE SCALES OF EFFECTIVE TEACHING



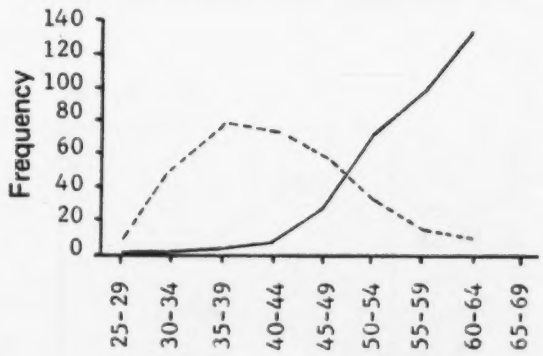
Scale 1. Analytic/Synthetic Approach



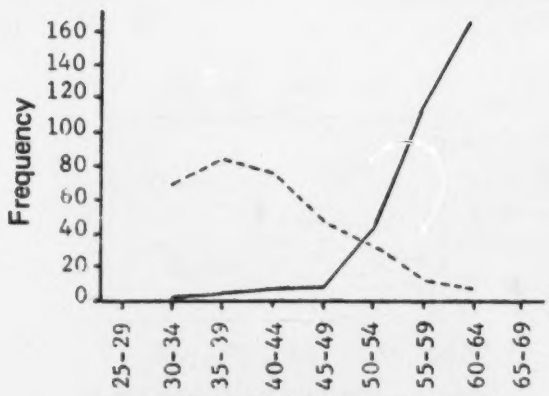
Scale 2. Organization/Clarity



Scale 3. Instructor-Group Interaction



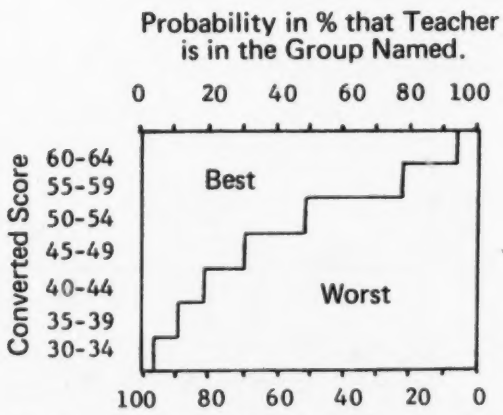
Scale 4. Instructor-Individual Student Interaction



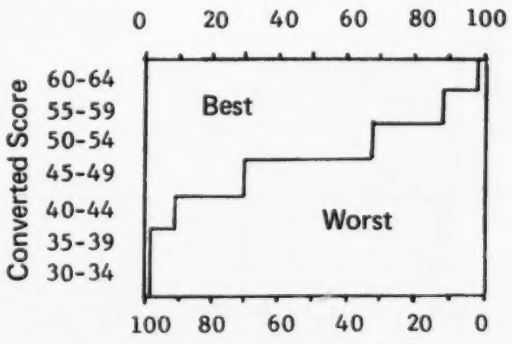
Scale 5. Enthusiasm/Dynamism

TABLE 9

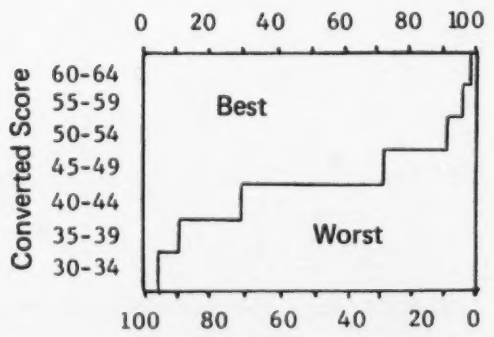
PROBABILITY CHARTS OF CONVERTED SCORES ($\bar{x}$ = 50, s = 10) OF 338 *BEST* AND 338 *WORST* TEACHERS ON FIVE SCALES OF EFFECTIVE TEACHING



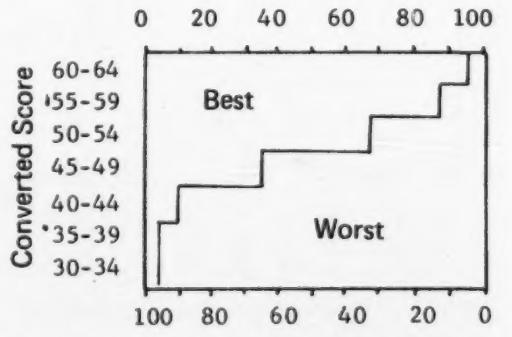
Scale 1. Analytic/Synthetic Approach



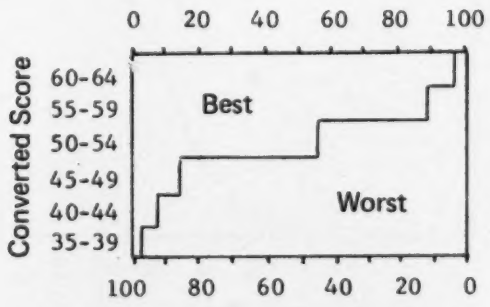
Scale 2. Organization/Clarity



Scale 3. Instructor-Group Interaction



Scale 4. Instructor-Individual Student Interaction



Scale 5. Enthusiasm/Dynamism

The scales are stressed because they have greater utility and conceptual value than the individual items. Even so, they do not include all of the useful data; some discriminating items do not cluster sufficiently with others to fall in any scale. A short evaluation form might well supplement the five summary descriptions with selections from items of this kind (for example, items from Table 3 that do not also appear in Table 5).

RATINGS OF TEACHERS RELATED TO CHARACTERISTICS OF COURSES AND STUDENTS

Courses and Students

To discover what variables significantly affect student ratings of teachers, the overall ratings of effectiveness of teaching from the 1968 survey were correlated with academic rank of teacher, course level, number of courses previously taken in the same department, class size, whether the course was required or optional, and whether the course was in the student's major or not. The highest correlation of any of these six variables with rated quality of teaching was .06, which is negligible. However, since the samples were large (N = 1015) for all variables except academic rank, course level, and class size (for which N = 51), statistical significance was achieved with a very small correlation; correlations bordering on the .05 level of significance were found for the last two variables listed. While these data confirm Solomon's (1966) data with respect to class size and Guthrie's (1954) results with respect to academic rank, they are partly in disagreement with a survey of class size at the University of Illinois noted by Cohen and Brawer (1969).

Although the six variables listed above are seen as not significantly influencing overall ratings of teaching effectiveness, they might be expected to be related to the scores assigned to teachers for each of the five student description scales of components of

effective teaching. Of the 30 elements of the matrix, only five coefficients are high enough (± .20 to ± .30) to establish a definite but small correlation: Scale 4, *Instructor-Individual Student Interaction*, correlates positively with higher level of course, smaller class size, and the course being in the major; Scale 1, *Analytic/Synthetic Approach*, correlates positively with higher level of course; and Scale 3, *Instructor-Group Interaction*, correlates positively with smaller class size. For 18 elements of the matrix, $p < .01$.

Turning to variables related more directly to the student, the 1015 overall ratings of teachers were correlated with sex of student, class level of student, grade-point average, and expected grade in course. All correlations were negligible (highest coefficient .09), although female sex and high expected grade in course correlated positively with high rating at just below and above the .01 level of significance. Cohen and Brawer (1969) reported similar results. Other studies have reported a relationship between expected grade and rating of teacher (Stewart & Malpass, 1966; Weaver, 1960), a relationship only at lower class levels (Anikeeff, 1953), and no relationship (Kent, 1967; Voeks & French, 1960). These contradictions seem consistent with the presence of a definite but trifling correlation.

The four variables listed above were also correlated with scores for each of the five student description scales of effective teaching. Of the 20 elements of the matrix, only one coefficient is high enough (.24) to be considered definite though small: Scale 4, *Instructor-Individual Student Interaction*, correlates positively with higher class level of student. Half of the correlations are significant at the .01 level or better. The matrix indicates that high achievers and advanced students are slightly less dependent than other students on organization and motivation supplied by the instructor, and also

that female students respond slightly more than males to personal and group interaction with their (predominantly male) instructors. Other investigators have related grade-point average to the needs, responses, and motivation of students (Downie, 1952; Spaights, 1967). The effects of authoritarianism, personality, and sex-related needs also have been studied (Carpenter et al., 1965; Doty, 1967; Freehill, 1967; McKeachie, 1963; Maney, 1959; and Rezler, 1965).

These results show that in general, the 10 course and student characteristics listed do not markedly affect student ratings of teachers. Measuring is usually not needed for these variables, and they might well be omitted from short evaluation forms. However, ratings of teachers having particular attributes may be somewhat influenced by certain of these variables (the personality of a particular teacher, for example, might tend to antagonize students of one sex more than the other). Analysis of the influence of course and student characteristics on ratings of teachers may, therefore, help individual instructors to adapt to local circumstance.

Two other relationships proved to be more marked. When number of nominations for most and least effective teachers ($N = 676$) were compared by subject areas, allowances being made for the sizes of the areas, differences significant at the .01 level were found. Corresponding analyses by type of course presentation revealed proportionately more *best* teachers in seminar courses than in lecture courses ($p < .001$), with lecture-laboratory courses being intermediate.

Goals of Students

Since effective teaching cannot be adequately understood without attention to the goals, perceptions, and values of students, these factors were studied in several ways.

TABLE 10

COLLEGE GOALS OF STUDENTS (N = 1015)

Scale 1. Upward Mobility/Security	Factor coefficient
1. To get the respect a college education brings	.72
2. To prepare for a better-paying job	.67
3. To earn a living more easily	.66
4. To gain greater security	.63
5. To have a better life than my parents	.50
6. To become a better citizen	.50
7. To associate with the preferred kind of people	.49
Scale 2. Self-Knowledge/Humanism	
8. To meet and learn from interesting people	.78
9. To learn more about myself and others	.75
10. To become more creative	.68
11. To broaden my overall viewpoint	.66
12. To be able to lead an interesting life	.45
Scale 3. Career/Subject Mastery	
13. To get the training needed for success	.83
14. To learn the skills needed for my career	.77
[illegible] To gain mastery of my field	.76
16. To earn the degree needed for my work	.60
17. To prepare for graduate school	.45

The 1967 student survey included 24 items on reasons for going to college. Responses were subjected to factor analysis and, following the procedures described above in the section on components of effective teaching, the results were validated in 1968. A three-scale solution having alpha reliabilities of .80, .81, and .81 was selected. Table 10 presents the scales and the 17 contained items with acceptable factor coefficients. Interpretations of the scales are indicated by the headings: Scale 1, *Upward Mobility/Security*; Scale 2, *Self-Knowledge/Humanism*; and Scale 3, *Career/Subject Mastery*. Items that did not appear in the scales tend

to relate to social pressure or apathy. Scale 1 has a low correlation (.30) with Scale 3, and the other intercorrelations are negligible.

Twenty items on students' perceptions of desirable objectives of teaching were processed into two scales having alpha reliabilities of .83 and .84 (Table 11). The interscale correlation is .01.

TABLE 11

OBJECTIVES OF TEACHING FAVORED BY STUDENTS (N = 1015)

Scale 1. Contribution to General Development	Factor coefficient
1. To help students mature	.73
2. To help students understand themselves	.68
3. To help students understand other people	.68
4. To help students develop their creative abilities	.66
5. To help students discover and develop their abilities	.65
6. To help students analyze their opinions and actions	.64
7. To teach students to communicate	.55
Scale 2. Transmission of Fundamentals	
8. To teach facts	.79
9. To teach fundamental principles	.74
10. To explain technical terms	.69
11. To transmit information	.65
12. To summarize important concepts	.60
13. To train students in the skills needed for their careers	.52

Relating the scales on college goals with those on objectives of teaching, *Contribution to General Development* has a somewhat moderate correlation with *Self-Knowledge/Humanism* (coefficient .54). *Transmission of Fundamentals* has moderate correlation with *Career/Subject Mastery* and low correlation with *Upward Mobility/Security* (coefficients .47 and .34, respectively).

Respondents to the 1968 survey were asked to rate their teachers, on a seven-point continuum, on constructive contributions made to their lives in each of six areas. Table 12 shows correlations of the mean scores for these areas with mean scores for the components of effective teaching and overall ratings of effectiveness of teaching.

TABLE 12

CORRELATIONS OF MEAN RATINGS OF CONSTRUCTIVE CONTRIBUTIONS WITH MEAN SCORES ON SCALES OF EFFECTIVE TEACHING AND OVERALL EFFECTIVENESS OF 51 TEACHERS

	Constructive Contribution to Life					
Components of effective teaching	Knowledge imparted	Counsel given	Objectives clarified	Values developed	Incentive elicited	Skills developed
1. Analytic/Synthetic Approach	.54	.28	.59	.50	.68	.14
2. Organization/Clarity	*.71*	.32	.49	.38	.58	.42
3. Instructor-Group Interaction	.44	.66	.49	.48	.66	.53
4. Instructor-Individual Student Interaction	.39	*.80*	.42	.46	.63	.45
5. Dynamism/Enthusiasm	.61	.38	.48	*.72*	*.74*	.33
Overall rating of effectiveness	*.79*	.55	.64	.62	.84	.55

Correlations > .70 = high (italicized); .70 to .40 = moderate; .40 to .20 = low. N = 51

Matching Students with Teachers

Correlations of both college goals and objectives of teaching with the components of effective teaching were low, with coefficients of the 25 elements ranging, for an N of 338, from -.19 to +.22. This doubtless results in part from the fact that only ratings of *best* teachers were utilized in the calculations. These teachers rated so high on all components of effective teaching that students with any goals and objectives can find in them attributes they admire; nevertheless, nine types of effective teachers were identified by analyzing individual patterns of relatively high and low scores on the five components of effective teaching. Overall ratings of teachers having the various patterns were then correlated with course and student variables. Because the analysis was complicated by many factors, results are not presented in numerical form lest the conclusions seem more exact than in fact they can be. The following two contrasting pairs of relationships are reported, however, to illustrate the concept of matching students with teachers.

Best teachers who were rated relatively high on Scale 4, *Instructor-Individual Student Interaction,* tended to be giving small lecture-laboratory classes, were particularly favored by female, upper-division and graduate students with low *Upward Mobility/Security* who valued *Contribution to General Development* and majored in the arts. By contrast, teachers who were rated relatively low on the same scale tended to be giving large lecture classes, were particularly favored by female and lower-division students with moderate *Upward Mobility/Security* who valued the *Transmission of Fundamentals.*

Best teachers who were rated relatively high on Scale 2, *Organization/Clarity* tended to be giving large lecture or lecture-laboratory classes, were particularly favored by male, lower-division students with high *Upward Mobility/Security* who

valued the *Transmission of Fundamentals* and majored in the biological sciences. By contrast, teachers who were rated relatively low on the same scale tended to give lecture classes of various sizes, were particularly favored by female, senior students who valued *Self-Knowledge/Humanism* and *Contribution to General Development* and majored in the humanities.

It seemed probable that *controversial* teachers (rated excellent by some observers and poor by others) would be less even in their performance than *best* teachers: Some students might accept relatively poor performance in a given component, whereas others, with different goals and objectives, might not. To test this hypothesis, the within-individual variances between the converted (standardized) scores for each component of effective teaching and the mean converted score for all five components were calculated separately for 112 ratings of 32 *best* teachers and contrasted with those for 154 ratings of 48 *controversial* teachers. As predicted, the within-individual variances were greater for the latter group ($p < .01$), indicating that ratings of *controversial* teachers on the five components of teaching were more variable than they were for *best* teachers. This explains, in part, their controversial status when rated by students with various goals, and indicates that it might be well for such teachers to be matched with students who are most inclined to value their particular assets. These analyses did not, however, test specifically for the values which might account for the varying student judgments.

EVALUATIONS DISCUSSED

What is Effective Teaching?

Many consider teaching to be excellent in proportion to progress made by learners toward stated educational objectives

(Kent, 1967; McKeachie, 1963). However, while this concept is generally sound, it is difficult to apply to the characterization or evaluation of university teaching because there is now insufficient agreement either on objectives, or on who should determine them. And even if there were widely accepted specific objectives, it is unlikely that there could now be agreement on how to test progress toward the attainment of many of them. Facts learned from teachers can be tested, but their value cannot; the contribution a teacher makes to spiritual or emotional maturation cannot easily be assessed.

Another way to assess teaching would be to consider it excellent in proportion to its constructive contribution to the life of the learner. Such a contribution might be knowledge imparted, wisdom instilled, experience offered, counsel given, objectives clarified, human values developed, incentive and inspiration elicited, or skills developed. Effective teaching usually contributes to the life of the student in several ways according to the individual teacher-student relationship. The learner may not be able to fully assess the constructive contribution made to his life by a teacher, and his judgment may change with time. Nevertheless, the learner is often (or usually) the best judge of contributions made to his own life. For this concept of effective teaching to be generally applicable, different students must tend to judge the same teachers as having made constructive contributions to them. This study indicates that in fact they do.

No definition of effective teaching was included in the questionnaires, leaving it to each respondent to select *best* and *worst* teachers by his own criteria. A descriptive definition of good teaching as actually perceived by students and colleagues was thus derived (Tables 3, 4, 5, and 6). The uniformity of judgment found in both the identification of *best* and *worst* teachers, and in the

characterization of *best* teaching, makes it clear that this descriptive approach is both practical and generally consistent with both of the views discussed above of how good teaching can best be assessed.

Other opinions, not seriously considered in this study, are that teaching should be judged primarily by students' increased ability to solve assigned problems (Beichl, 1967), by out-of-class accomplishments (Brandis, 1964), or by the academic prowess of former students.

Comparison of Evaluations by Students and Colleagues

Colleague Scales 1, *Research Activity and Recognition,* and 2, *Intellectual Breadth*, relate to scholarship as expressed in research. Excellence in research is clearly not sufficient ground for establishing excellence in teaching, particularly at the undergraduate level, and it is highly inappropriate that at most institutions research productivity is the primary consideration in evaluating teaching ability (Astin & Lee, 1966).

Colleagues tended to rate full professors relatively high on Scale 1, doubtless because it takes time to establish a reputation for competence in research, even though professorial rank as such did not affect student or faculty ratings of teaching. "Professional competence" also is a criterion for advancement at many universities, but since measures of professional competence (e.g., positions held, honors received) are largely responses to reputation beyond the home campus for research rather than teaching, research is, in effect, counted another time. Therefore, when excellence in research is considered separately as a criterion for advancement, it should specifically be eliminated in evaluating effectiveness of teaching; Colleague Scale 1, and items 10 and 13 of Scale 2, should not be used for rating teaching. Student Scale 1, *Analytic/Synthetic*

Approach, is not equivalent to Colleague Scales 1 and 2, but does also relate to scholarship; if this scale is used, scholarship would be considered as it is expressed in teaching.

Colleague Scale 3, *Participation in the Academic Community*, appears to be relatively weak conceptually, although the items composing the scale are individually satisfactory. Ratings of teachers made by the various members of the academic community are rarely completely independent: Communication between students and between faculty and students influences judgments. This is particularly true for the information elicted from items in Colleague Scale 4, *Relations with Students*, which faculty members usually get indirectly, from students' comments. Accordingly, Colleague Scale 4 appears to us to be less direct, more superficial, and hence less valid than the related Student Scales 3, *Instructor-Group Interaction*, and 4, *Instructor-Individual Student Interaction*.

Items 30 through 45 of the colleague survey (Table 4) relate to teaching observed in seminars and in the classroom. However, 17 percent of the faculty respondents had not attended a seminar given by the teacher they had selected as *best*, 51 percent had not observed classroom teaching of the teacher they considered *best*, and a surprising 75 percent had not observed classroom teaching of the teacher they thought was *worst*. Further, most members of the faculty who had observed the teaching of the named colleague had done so only briefly or infrequently.

We conclude that ratings by colleagues should be used to supplement, though not to substitute for, ratings by students; accordingly, our analysis stresses the student scales. However, Colleague Scale 5 *Concern for Teaching* relates directly to teaching and is based on items that faculty, not students, can observe. This

scale could profitably be represented in any evaluations of teaching made by colleagues.

Discussion of the ways in which both colleagues and students may provide environmental encouragement for effective teaching can be found in Gaff and Wilson (in press).

Sample Size and Norms

It is essential that teacher evaluations be based on adequate samples of opinion. About 25 responses might be considered minimal, and a return rate of at least 50 percent is desirable. Teachers regarded as excellent by some observers and poor by others should be rated by as many observers as possible. Teachers of even small classes can be rated adequately if an acceptable number of evaluations are accumulated over time.

Whether the teaching of individuals and departments should be evaluated on an absolute or relative basis is open to question. In practice, however, academic advancement, and students' choices of courses and curricula, are often based at least in part on comparisons of teacher with teacher and department with department. It is important, therefore, that norms be established so that scores can be interpreted. Norms should be calculated at the campus level for some elements of any evaluation form used in promotion procedures, and the summary descriptions of the five principal components of effective teaching would be satisfactory for the calculation of such norms. Departments or subject areas might find it useful also to calculate their own norms, particularly if they have developed their own evaluation forms, but it is desirable that any norms used be recalculated at frequent intervals to assure that the system of evaluation is being responsive to change.

A Potential Weakness in the Use of Student Evaluations

It is unlikely that an instructor could use the findings of this study to elicit higher student ratings than he deserves; scholarship, rapport, and enthusiasm are difficult to simulate, and students are not easily deceived. There are circumstances, however, which can adversely affect a good teacher's performance: His work load may be too heavy, his classes may be too large, he may have been assigned to teach outside the area of his greatest competence, his course may be new and untried, or he may be experimenting with innovations. And although the student properly rates his teacher on how good he perceives the instruction to be, not on how good it could have been or will become, it would be unfortunate if rating procedures either penalized teachers for factors beyond their control or encouraged them to offer only "safe," familiar instruction. This danger can be minimized if it is recognized and appropriate steps taken to bring any such special circumstances to the attention of the administration.

The instructor might be granted the option of retaining ratings for his exclusive use the first time his teaching of any one course is evaluated. Thereafter all returns should be transmitted at least to the department, but we suggest that provisions be made so that the instructor can challenge individual returns that seem malicious or invalid, and file a comment on the ratings if he so wishes.

Alternative Student Evaluation Forms

The results of this study can be used in many ways, depending on objectives and facilities. Three kinds of evaluations, intended to be suggestive rather than limiting, are discussed below:

Long form. The 85 items of Table 3 provide the basis for a long evaluation form. The list might well be altered to better adapt it to the requirements of a particular teacher, department, or subject area. Use of such a form provides much information and thus is useful to teachers, whether new or established, who wish to improve. (Some instructors believe, that a single open-ended question such as, *You are invited to comment further on the course and/or effectiveness of the instructor,* elicits the most useful responses for this purpose.) A long form, however, is relatively slow to complete, and results, being diverse, would be difficult to apply to advancement procedures. This being true, evaluations would probably be ignored by some teachers.

Short form. The basis for a short evaluation form is provided by the five summary descriptions of the components of effective teaching, (see p.18) supplemented by additional discriminating items not represented in those scales (for example, items 9, 15, 20, 21, 23, 24, 40, 46, 55, and 79 of Table 3). Such an instrument would be effective for evaluating teaching for use in advancement procedures. It is applicable to most university teaching and therefore would permit the calculation of departmental, college, and campus norms. A short form is less directly useful than a long form for helping teachers to improve their performance, although it is highly probable that if teaching were to become a more effective criterion for academic advancement, performance would improve.

Medium-length form. An evaluation form of medium length might provide a desirable compromise between the advantages and disadvantages of longer and shorter forms. The 36 items of Table 5 supplemented by the same 10 items cited for the short form would be satisfactory. Some demographic items also might be included.

THE PRINCIPAL RESULTS

I. Analysis of the items characterizing *best* teachers as perceived by students produced five scales, or components of effective performance (Table 5). The conceptual interpretations of the scales are indicated by the headings assigned:

1. *Analytic/Synthetic Approach*
2. *Organization/Clarity*
3. *Instructor-Group Interaction*
4. *Instructor-Individual Student Interaction*
5. *Dynamism/Enthusiasm*

II. Analysis of the items characterizing *best* teachers as perceived by colleagues produced five scales (Table 6):

1. *Research Activity and Recognition*
2. *Intellectual Breadth*
3. *Participation in the Academic Community*
4. *Relations with Students*
5. *Concern for Teaching*

III. Eighty-five items are listed that characterize *best* teachers as perceived by students (Table 3), and 54 items are listed that characterize *best* teachers as perceived by colleagues (Table 4). All items statistically discriminate *best* from *worst* teachers with a high level of significance.

IV. The student scales were derived from a 1967 survey. A single summary description was phrased to express the nature of the component of effective teaching identified by the items composing each scale. Respondents to the 1968 survey rated their teachers on each of the five summary descriptions and also on each

of the items from which the scales had been derived. Correlations of mean scores on the summary descriptions with mean scores on the full lists of respective items were very high. Thus, the five summary descriptions provide the basis for a short evaluation form demonstrated to be broad and highly discriminating.

V. In general, student ratings of *best* teachers showed only negligible correlations with academic rank of instructor, class level, number of courses previously taken in the same department, class size, required versus optional course, course in major or not, sex of respondent, class level of respondent, grade-point average, and expected grade in course.

VI. There is excellent agreement among students, and between faculty and students, about the effectiveness of given teachers.

VII. *Best* and *worst* teachers engage in the same professional activities and allocate their time among academic pursuits in about the same ways. The mere performance of activities associated with teaching does not assure that the instruction is effective.

VIII. A disproportionate number of *best* teachers were teaching seminar rather than lecture courses, and a wide range of excellence was revealed in the teaching of different subject areas.

IX. Analysis of 17 items describing the college goals of students produced three scales (Table 10):

1. *Upward Mobility/Security*
2. *Self-Knowledge/Humanism*
3. *Career/Subject Mastery*

X. Analysis of 13 items describing objectives of teaching as perceived by students produced two scales (Table 11):

1. *Contribution to General Development*
2. *Transmission of Fundamentals*

XI. Students evaluated the positive contributions made to their lives by *best* teachers in six areas: knowledge imparted, counsel given, objectives clarified, values developed, incentive elicted, and skills developed. Correlations of mean scores for these areas with mean scores for the components of effective teaching and with overall ratings of effectiveness of teaching are high (Table 12).

XII. Nine types of effective teachers were identified by analyzing individual patterns of relatively high and low scores on the five components of effective teaching. Overall ratings of teachers having the various patterns correlate with certain course and student variables.

XIII. Teachers rated as excellent by some observers and as poor by others are less even in their performance of the five components of effective teaching than are *best* teachers.

SOME IMPLICATIONS

The study has shown that different types of teaching appropriate to different settings can be assessed, that a variety of types of effective teaching can be identified, and that use of an evaluation instrument does not presume that there is only one type of effective teaching–that it is possible, in short, to develop procedures for the systematic evaluation of college and university teaching.*

*One of the authors (Hildebrand, 1971) has responded elsewhere to objections that are commonly raised to the use of students' evaluations of teaching.

As the concern for improving the quality of teaching mounts, and the critical importance of teaching in the lives of students is increasingly recognized, the basic question that has long been asked about teaching evaluations inevitably broadens. Since it is clear now that evaluation is continuous and inescapable on every campus, we can no longer afford to ask, "Should teaching be evaluated?" The question becomes, rather, "Do we have valid and systematic ways for eliciting the evaluations that are made?" The results of the present study speak to this larger issue and provide, through the instruments developed, a means for securing the necessary information from students and faculty. Such basic support from research is critical to the identification and encouragement of effective teaching.

Planning for Programs of Teacher Evaluation

The purposes and resources of individual colleges and universities vary, and the committees and individuals charged with teacher evaluation on particular campuses usually want to put their own unique imprints on whatever programs are used at their institutions. Because of this, a single prepackaged product for teacher evaluation is generally not acceptable.

Nevertheless, whatever the variations in local options, there are some key decisions which must be made in developing a successful program of teacher evaluation. It has become evident that the chief sources of disillusionment with programs for teacher evaluation arise from the failure to develop sufficiently detailed plans which spell out key decisions and anticipate realistic difficulties and possible controversies.

The following outline is intended to assist planners by spelling out a number of tasks to be undertaken and options to be considered in implementing an evaluation program.

PURPOSES

Feedback to instructor for self-improvement
Data for making salary, promotion, and tenure decisions
Information to assist students in choosing courses and instructors
A combination of the above

SCOPE

Number of Teachers

Small number (e.g., all of one department)
Medium number (e.g., all eligible for tenure)
Large number (e.g., all in the institution)

Number of classes

One per instructor per advancement period
One per instructor per year
Each once per advancement period
Each every other year, or every year

Number of students

Random sample of X students (large classes only?)
X percent of class (large classes only?)
All (but with minimum of X returns to qualify for interpretation?)

Kinds of courses

Undergraduate credit courses
All except seminars and field research courses
All (including noncredit and extension?)

FORMS

Style

Structured check-off items
Open-ended essay items

Coverage

Teaching only
Teaching and course
Teaching, course, and student data (demographic, objectives, values)

Format

Optical scanning sheets
Mark sense sheets

Porto-punch cards
Duplicated questionnaire with key punch
Duplicated questionnaire with hand tally

Length

Short (1-25 items)
Medium (26-50 items)
Long (> 50 items)

Sources

External (for example, another campus, Center for Research and Development in Higher Education, Berkeley)
Local committee (faculty, administrative, student, combination)
Instructor
A combination of the above

ADMINISTRATION AND DATA GATHERING

Time of distribution

Early in course
Late in course
With final examination
After course

Method of distribution

Instructor
Student representative
Administrative representative
With registration packets
Mail

Method of return

Collected by instructor
Collected by student representative
Collected by administrative representative
Mailed to a central office

DATA REDUCTION

Persons involved

Instructor
Department
Committee (student, faculty, administrative, combination)
Central office

Method

Summarization by computer, with norms and variances
Hand-tabulation and individual case study
Summarization of open-ended data

INTERPRETATION OF DATA

Persons involved

Instructor
Department
Committee (student, faculty, administrative, combination)
Central office

Basis

Individual case study
Departmental norms
College or school norms
Campus norms

PROVISION FOR CHALLENGE

None
By instructor
By students or department
Procedures

DISSEMINATION AND REPORTING

To instructor only
To instructor and departmental chairman or committee
To instructor, department, and administration
To university community at a central location
To university community by sale or general distribution

References

Anikeeff, A. M. Factors affecting student evaluation of college faculty members. *Journal of Applied Psychology*, 1953, 37, 458-460.

Astin, A. W. & Lee, C. B. T. Current practices in the evaluation and training of college teachers. *The Educational Record*, 1966, 47, 361-375.

Beichl, G. J. The genesis of creativity. *Science*, February 17, 1967, 155, p.774.

Bendig, A. W. An inverted factor analysis study of student-rated introductory psychology instructors. *Journal of Experimental Education*, 1953, 21, 333-336.

Brandis, R. The rehabilitation of university undergraduate teaching. *The Educational Record*, 1964, 45, 56-63.

Carpenter, F., Van Egmond, E., & Jochem, J. Student preference of instructor types as a function of subject matter. *Science Education*, 1965, 49, 235-238.

Coffman, W. E. Determining students' concepts of effective teaching from their ratings of instructors. *Journal of Educational Psychology*, 1954, 45, 277-286.

Cohen, A. M., & Brawer, F. B. *Measuring faculty performance.*

Washington, D. C.: American Association of Junior Colleges, 1969.

Cosgrove, D. J. Diagnostic rating of teacher performance. *Journal of Educational Psychology*, 1959, 50, 200-204.

Crannell, C. W. A preliminary attempt to identify the factors in student-instructor evaluation. *Journal of Psychology*, 1953, 36, 417-422.

Doty, B. A. Teaching method effectiveness in relation to certain student characteristics. *Journal of Educational Research*, 1967, 60, 363-365.

Downie, N. M. Student evaluation of faculty. *Journal of Higher Education*, 1952, 23, 495-496, 503.

Estrin, H. A. Effective and ineffective engineering instructors. *Improving College and University Teaching*, 1965, 13, 137-140.

Freehill, M. F. Authoritarian bias and evaluation of college experiences. *Improving College and University Teaching*, 1967, 15, 18-19.

French, G. J. *College students' concept of effective teaching determined by an analysis of teacher ratings.* Seattle: University of Washington, 1957. Mimeographed.

Gaff, J. G., & Wilson, R. C. The teaching environment. *American Association of University Professors Bulletin*, in press.

Garverick, C. M., & Carter, H. D. Instructor ratings and expected grades. *California Journal of Educational Research*, 1962, 13, 218-221.

Gibb, C. A. Classroom behavior of the college teacher. *Educational and Psychological Measurement*, 1955, 15, 254-263.

Gowin, D. B., & Payne, D. E. Evaluating instruction: Cross-perceptions of college students and teachers. *School Review*, 1962, 70, 207-219.

Gustad, J. W. Policies & practices in faculty evaluation. *The Educational Record*, 1961, 42, 194-211.

Guthrie, E. R. *The evaluation of teaching: A progress report.* Seattle: University of Washington, 1954 (lithoprint).

Hayes, R. B. Ways to measure teaching. *Journal of Educational Research*, 1963, 57, 47-50.

Hildebrand, M. How to recommend promotion for a mediocre teacher without actually lying. *Experiment and Innovation: New Directions in Education at the University of California*, 1971, 4, 1-21.

Hodgson, T. F. *The general and primary factors in student evaluation of teaching ability.* Seattle: University of Washington, 1958 Mimeographed.

Isaacson, R. L., McKeachie, W. J., Milholland, J. E., Lin, Y. G., Hofeller, M., Baerwaldt, J. W., & Zinn, K. L. Dimensions of student evaluations of teaching. *Journal of Educational Psychology*, 1964, 55, 344-351.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.

Kent, L. Student evaluation of teaching. In C. B. T. Lee (Ed.), *Improving college teaching*. Washington, D. C.: American Council on Education, 1967, Pp.312-343.

Lacognata, A. A. University extension faculty and student role expectations: An empirical analysis. *Journal of Experimental Education*, 1964, 33, 107-120.

Maney, A. C. The authoritarianism dimension in student evaluations of faculty. *Sociology of Education*, 1959, 32, 226-231.

McKeachie, W. J. Research on teaching at the college and university level. In N. L. Gage (Ed.), *Handbook of Research on Teaching*. Chicago: Rand McNally, 1963. Pp. 1118-1172.

Regan, M., & Yonge, C. D. Educational relevance: The meaning of college. Private communication: long-range study in progress.

Remmers, H. H., & Baker, P. C. *Manual of instructions for Purdue rating scale for instruction*. Lafayette, Ind.: Purdue University (Division of Educational Reference), 1952.

Rezler, A. G. The influence of needs upon the student's perception of his instructor. *Journal of Educational Research*, 1965, 58, 282-286.

Ryans, D. G. *Characteristics of teachers: Their description, comparison, and appraisal.* Washington, D. C.: American Council on Education, 1960.

Solomon, D. Teacher behavior dimensions, course characteristics, and student evaluations of teachers. *American Educational Research Journal*, 1966, 3, 35-47.

Solomon, D., Rosenberg, L., & Bezdek, W. E. Teacher behavior and student learning. *Journal of Educational Psychology*, 1964, 55, 23-30.

Spaights, E. Students appraise teachers' methods and attitudes. *Improving College and University Teaching*, 1967, 15, 15-17.

Stewart, C. T., & Malpass, L. F. Estimates of achievement and ratings of instructors. *Journal of Educational Research*, 1966, 59, 347-350.

Tryon, R. C., & Bailey, D. E. The BC TRY computer system of cluster and factor analysis. *Multivariate Behavioral Research*, 1966, 1(1), 95-111.

Voeks, V., & French, G. M. Are student ratings of teachers affected by grades? *Journal of Higher Education*, 1960, 31, 330-334.

Weaver, C. H. Instructor rating by college students. *Journal of Educational Psychology*, 1960, 51, 21-25.

Wherry, R. J. *Control of bias in rating: Instructor rating scales.* Washington, D. C.: Personnel Research Section, AGO, United States Department of the Army, 1950.

Wilson, R. C., Gaff, J. G., & Bavry, J. L. *Manual of information, faculty characteristics questionnaire (Experimental form I).* Berkeley: Center for Research and Development in Higher Education, 1970, Mimeographed.